The Channel Capacity of a Fiber Optics Communication System: perturbation theory

Evgenii Narimanov and Partha Mitra

Abstract— **We consider the communication channel given by a fiber optical transmission line. We develop a method to perturbatively calculate the information capacity of a nonlinear channel, given the corresponding evolution equation. Using this technique, we compute the decrease of the channel capacity to the leading order in the perturbative parameter for fiber optics communication systems.**

I. Introduction

THe performance of any communication system is ultimately limited by the signal to noise ratio of the received signal and available bandwidth. This limitation can be stated more formally by using the concept of *channel capacity* introduced withing the framework of information theory[1]. The channel capacity is defined as the maximum possible bit rate for error-free transmission in the presence of noise. For a linear communication channel with additive white Gaussian noise, and a total signal power constraint at the input, the capacity is given by the celebrated Shannon formula[1]

$$C \quad = \quad W \log\left(1 + \frac{P_0}{P_N}\right) \tag{1}$$

where W is the channel bandwidth, P_0 is the average signal power, and P_N is the average noise power.

Current optical fiber systems operate substantially below the fundamental limitation, imposed by the Eq. (1). However, a considerable improvement in the coding schemes for lightwave communications, expected in the near future, may result in the development of systems, whose efficiency may approach this fundamental limit.

However, the representation of the channel capacity in the standard form (1) is unsuitable for applications to the actual fiber optics systems. It was obtained based on the assumption of *linearity* of the communication channel, while the modern fiber optics systems operate in a substantially nonlinear regime. Since the optical transmission lines must satisfy very strict requirements for bit-error-rate (10^{-12} to 10^{-15}), the pulse amplitude should be large enough so that is can be effectively detectable. The increase of the number of wavelength-division multiplexing (WDM) channels[2] in the modern fiber optics communication systems also leads to a substantial increase of the electric field intensity in the fiber. As a consequence, the Kerr nonlinearity of the fiber refractive index $n = n_0 + \gamma I$ (where I is the pulse intensity) becomes substantial and should be taken into account.

In the present paper we consider corrections to the channel capacity of the optical fiber communication system, originating from the nonlinearity of the fiber. The technique that we use involves a perturbative computation of the relevant mutual information and subsequent optimization. To our knowledge, this method appears to be substantially new.

II. Fiber Optics Communication System as an Information Channel

We consider a typical fiber optics communication system, which consists of a sequence of N fibers each followed by an amplifier (see Fig. 1). The amplifiers have to be introduced in order to compensate for the power loss in the fiber. An inevitable consequence of such design, however, is the generation of the noise in the system, coming from the spontaneous emission in the optical amplifiers. For simplicity, we will assume that all the fibers and the amplifiers of the link are identical.

The information is encoded in the electric field at the "imput" of the system, typically using the light pulses sent at different frequencies. The available bandwidth of the amplifiers as well as the increase of the fiber absorption away from the "transparency window" near the wavelength $\lambda = 1.55\mu m$, limits the bandwidth of the fiber optic communication system.

The maximum amount of the information, that can be transmitted through the communication system per unit time, is called the channel capacity C. According to the Shannon's basic result[1], this quantity is given by the maximum value of the mutual information per second over all possible input distributions:

$$C \quad = \quad \max_{p_x}\left\{H\left[y\right] - \left\langle H\left[y \mid x\right]\right\rangle_{p_x}\right\} \tag{2}$$

The mutual information

$$R \quad = \quad H\left[y\left(\omega\right)\right] - \left\langle H\left[y\left(\omega\right)|x\left(\omega\right)\right]\right\rangle_{p_x} \tag{3}$$

is a functional of the "input distribution" $p_x\left[x\left(\omega\right)\right]$, which represents the encoding of the information using the electric field components at different frequences

$$E_{\rm in}\left(t\right) \quad = \quad \int_W d\omega\ x\left(\omega\right)\exp\left(i\omega t\right) \tag{4}$$

The entropy $H\left[y\left(\omega\right)\right]$ is the measure of the information received at the output of the communication channel. However, if the channel is noisy, for any output signal there is some uncertainty of what was originally sent. The conditional entropy $H\left[y\left(\omega\right)|x\left(\omega\right)\right]$ at the output *for a given* $x\left(\omega\right)$ represents this uncertainty.

The entropies $H\left[y\left(\omega\right)\right]$ and $H\left[y\left(\omega\right)|x\left(\omega\right)\right]$ are defined in terms of the corresponding distributions $p\left(y\right)$ and $p\left(y|x\right)$ via the standard relation

$$H \quad \equiv \quad -\int \mathcal{D}y\left(\omega\right) p\left[y\left(\omega\right)\right]\log\left(p\left[y\left(\omega\right)\right]\right) \tag{5}$$

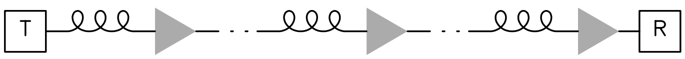

Fig. 1. The schematical representation of a fiber optics communication channel

where $p \equiv p_y(y)$ for the entropy $H[y(\omega)]$, and $p \equiv p(y \mid x)$ for the entropy $H[y(\omega)|x(\omega)]$, and the functional integral is defined in the standard way

$$\int \mathcal{D}\xi(\omega) \equiv \lim_{M\to\infty} c_M \left[\Pi_{m=1}^{M} \int d\xi(\omega_m)\right] \quad (6)$$

where is a normalization constant.

For any communication link, the signal power is limited by the system hardware. Therefore, the maximum of the mutual information in (1) should be found under the constraint of the fixed total power P_0 at the input:

$$P_0 = \int \mathcal{D}x(\omega) |x(\omega)|^2 p_x[x(\omega)] \quad (7)$$

If the propagation in the communication channel is described by a linear equation, then the input-output relation for the system is given by

$$y(\omega) = K(\omega) x(\omega) + n(\omega) \quad (8)$$

where $n(\omega)$ is the noise in the channel. In this approximation, the problem of finding the maximum of the mutual information can be solved exactly, with the corresponding input distribution p_x being Gaussian[1]. If the amplifiers compensate exactly for the power losses in the fibers, the Channel Capacity is given by the Shannon formula (1).

As follows from (1), the better bit rates can be obtained for the higher signal-to-noise ratio P_0/P_N. With this in mind, the optics fiber communication systems are designed to operate with the pulses of high power. As a result, the optics fiber links operate in the regime, in which due to the Kerr nonlinearity the refraction index of the fiber strongly depends on the local electric field intensity. Therefore, a modern fiber optics communication system is, in fact, an essentially nonlinear communication channel, and cannot be adequately described within the framework of the Shannon's *linear* theory.

III. The Model

The first step in the calculation of the channel capacity is to find the "input-output" relation for the communication channel. The time evolution of the electric field in the fiber $E(z,t)$, where z is the distance along the fiber, can be accuarately described in the "envelope approximation"[2], when

$$E(z,t) = A(z,t) \exp\left(i\left(\beta_0 z - \omega_0 t\right)\right) + \text{c.c.} \quad (9)$$

where the function A represents the slowly (compared to the light frequency) varying amplitude of the electric field in the fiber. The evolution of $A(z,t)$ is described by the equation

$$\frac{\partial A}{\partial z} + \beta_1 \frac{\partial A}{\partial t} + \frac{i}{2}\beta_2 \frac{\partial^2 A}{\partial t^2} + \frac{\alpha}{2} A = i\gamma |A|^2 A \quad (10)$$

Here the coefficients β describe the frequency dependence of the wavenumber

$$\beta(\omega) = \beta_0 + \beta_1 \omega + \frac{\beta_2}{2}\omega^2 + O\left[\omega^3\right] \quad (11)$$

where ω is measured from the center of the band ω_0.

The equation (10) neglects the effects such as the stimulated Raman scattering and the stimulated Brillouin scattering[2], compared to the Kerr nonlinearity of the refraction index of the fiber, represented by the term $\gamma |A|^2 A$.

The optical amplifiers incorporated into the communication system (see Fig. 1) compensate for the power losses in the fiber, but due to spontaneous emission each of them will inevitably introduce noise $n(t) = \exp(i\omega_0 t) \int d\omega \ n_\omega \exp(i\omega t)$ into the channel. Generally, even in a single optical amplifier, the noise distribution at any given frequency $\omega_0 + \omega$ within the channel bandwidth $n(\omega)$ is close to a Gaussian:

$$p_n[n(\omega)] \sim \exp\left[-\frac{|n(\omega)|^2}{P_N^\omega}\right] \quad (12)$$

This is even more so in a system with many independent amplifiers, due to the Central Limit Theorem. For simplicity, the noise spectrum P_N^ω can be assumed to be flat.

If the envelope function just before the amplifier is $A(t) \equiv A_\omega^0 \exp(i\omega t)$, then immediately after the amplifier

$$A_\omega = \exp\left(\frac{\alpha}{2} d\right) A_\omega^0 + n_\omega \quad (13)$$

where d is the span of a single fiber.

The equations (10), (13) define the evolution of the electric field envelope over one "fiber-aplifier" link of the communication system. The total "input-output" relation will then involve solving the corresponding equations for all N iterations of the single fiber-amplifier unit.

IV. The Perturbative Framework

If one is able to calculate the "output" signal $y(\omega)$ in terms of the "input" $x(\omega)$ and the noise contributions of each of the amplifiers $n_\omega^{\{\alpha\}}, \alpha = 1, \ldots, N$,

$$y(\omega) = \Phi\left(x(\omega); n_\omega^{\{1\}}, \ldots, n_\omega^{\{N\}}\right) \quad (14)$$

then the the conditional distribution $p(y|x)$ can be simply calculated as follows:

$$\begin{aligned} p(y|x) &= \left\{\Pi_{\alpha=1}^{N-1} \int \mathcal{D}n_\omega^{\{\alpha\}} \ p_n\left[n_\omega^{\{\alpha\}}\right]\right\} \\ &\times \ p_n\left[y(\omega) - \Phi\left(x(\omega); n_\omega^{\{1\}}, \ldots, n_\omega^{\{N\}}\right)\right] \end{aligned} \quad (15)$$

where p_n is the distribution function of the noise, produced by a single amplifier. The output distribution $p_y(y)$ can then be directly related to the input distribution $p_x(x)$ via the standard relation

$$p_y(y) = \int \mathcal{D}x(\omega)\ p[y(\omega)|x(\omega)]\ p_x[x(\omega)] \tag{16}$$

Using Eqns. (15),(16), one is able to express the mutual information in terms of a single distribution p_x. The calculation of the channel capacity then reduces to a standard problem of finding the maximum of a (nontrivial) functional.

The equation (10) is, in fact, the well studied [3] nonlinear Shroedinger equation, with the time and distance variables interchanged. Only some partial solutions of this equation are known, corresponding to solitons[3], [4]. However, in order to calculate the channel capacity, one needs to find the general input-output relation for the communication system. This implies solving a set of N essentially nonlinear equations (10) for *arbitrary* initial conditions. Even knowing some partial solutions, doing such calculation exactly for an essentially nonlinear system is not possible in a closed form.

In order to make progress, we note the presence of a natural perturbation parameter in the problem, namely γ.In fact, the fiber equation (10) is already an approximation, derived in the limit, when the change in the effective refraction index due to pulse propagation, described by the nonlinear term $i\gamma|A|^2 A$, is *small* compared to the "unperturbed" value of the index of refraction n_0. We have developed a perturbative technique, when the solution of the nonlinear evolution equation, is represented as a power series in γ. Solving (10) separately for each power of γ, and using (13), for the input-output relation of a single fiber-amplifier unit $\Phi_\omega^{(n)}$, defined as

$$A_\omega^{(n)} = \Phi_\omega^{(n)}\left(A_\omega^{(n-1)}\right), \tag{17}$$

we obtain:

$$\begin{aligned}\Phi_\omega^{(n)}\left(A_\omega{}^{(n-1)}\right) &= \left[A_\omega^{(n-1)} + \sum_{\ell=1}^{\infty}\gamma^\ell \mathcal{F}_\omega^{(\ell)}\left(A_\omega^{(n-1)}\right)\right] \\ &\times\ \exp\left(-i\kappa_\omega d\right) + n_\omega \end{aligned} \tag{18}$$

where d is the length of a single fiber,

$$\kappa_\omega = \beta_1\omega - \frac{1}{2}\beta_2\omega^2 \tag{19}$$

The procedure for the calculation of the functions $\mathcal{F}_\omega^{(\ell)}$, described in detail Appendix A, can be carried to an arbitrary order ℓ.

The further calculation then involves the following steps:

- Iterating Eq. (18) N times, to obtain the "input-output" relation for the whole communication system $\Phi_\omega\left[x(\omega); n_\omega^{(1)}, \ldots, n_\omega^{(N)}\right]$
- substituting the result into Eqns. (15), (16) to obtain the conditional distribution $p(x|y)$ and the output distribution $p_y(y)$ in terms of the input distribution $p_x(x)$ as expansions in powers of γ
- calculating the entropies $H[y(\omega)]$ and $H[x(\omega)|y(\omega)]$, and the mutual information R

Following these steps, the calculation of the channel capacity becomes a straightforward procedure. In Appendix B we describe it in detail, using a simple nonlinear channel $y(\omega) = x(\omega)\exp(-\phi[x(\omega)]) + n(\omega)$ as an example.

V. The Fiber Link Channel Capacity

After a tedious, but straigthforward calculation, we obtain:

$$H[y(\omega)] = H_0[y(\omega)] - \Delta C_1 - \Delta H_y + \mathcal{O}\left(\gamma^4\right) \tag{20}$$

and

$$H[y(\omega)|x(\omega)] = H_0[y(\omega)|x(\omega)] + \Delta C_2 + \mathcal{O}\left(\gamma^4\right) \tag{21}$$

where $H_0[y]$ and $H_0[x|y]$ are given by the standard expressions for a linear channel[1]. In the limit of large signal-to-noise ratio $P_0 \gg P_N$ we obtain:

$$\Delta C_1 = N^2 W\left(\frac{\gamma P_0}{\alpha}\right)^2 Q_1\left(\alpha d, \frac{\beta_2^2 W^4}{\alpha^2}\right) \tag{22}$$

$$\Delta C_2 = \frac{4}{3}\left(N^2-1\right) W\left(\frac{\gamma P_0}{\alpha}\right)^2 Q_2\left(\alpha d, \frac{\beta_2^2 W^4}{\alpha^2}\right) \tag{23}$$

and the functions Q_1 and Q_2 are defined as follows:

$$\begin{aligned}Q_1(u,z) &= \int_{-1/2}^{1/2} dx_1 \int_{-1/2}^{1/2} dx_2 \int_{1/2}^{\bar{x}} dx \\ &\times\ f(u,z;x_1,x_2,x) \end{aligned} \tag{24}$$

$$\begin{aligned}Q_2(u,z) &= \int_{-1/2}^{1/2} dx_1 \int_{-1/2}^{1/2} dx_2 \int_{-1/2}^{1/2} dx \\ &\times\ f(u,z;x_1,x_2,x) \end{aligned} \tag{25}$$

where $\bar{x} \equiv \max[1/2, 1/2 + x_1 + x_2]$, and

$$f(u,z;x_1,x_2,x) \equiv \frac{\left|1-\exp\left(-u-iv^2\right)\right|^2}{1+v^2} \tag{26}$$

where

$$v \equiv z(x-x_1)(x-x_2) \tag{27}$$

The correction

$$\Delta H_y = \gamma^2 W \int \mathcal{D}y(\omega) p_y^{(0)}[y(\omega)]\left(p_y^{(1)}[y(\omega)]\right)^2 \tag{28}$$

is caused by the deviations of the otput distribution

$$p_y = p_y^{(0)} + \sum_\ell \gamma^\ell p_y^{(\ell)} \tag{29}$$

from the Gaussian form

$$p_y^{(0)}[y(\omega)] \sim \exp\left(-\frac{|y(\omega)|^2}{P_\omega + P_\omega^N}\right) \tag{30}$$

where P_ω (such that $P_0 = \int d\omega P_\omega$) is the input power at frequency ω:

$$p_x^{(0)}\left[x\left(\omega\right)\right] \sim \exp\left(-\frac{\left|x\left(\omega\right)\right|^2}{P_\omega^0}\right) \quad (31)$$

Note, that the correction $\Delta H_y \geq 0$ and equals to zero only when the distribution $p_y^{(1)} = 0$. Therefore, as follows from Eqns. (20),(28), in the second order in nonlinearity γ the mutual information R has the maximum, when $p_y^{(1)} = 0$, or, equivalently, when the *output* distribution is Gaussian up to the *first* order in nonlinearity.

For a general nonlinear channel, that would correspond to the input distribution, being non-Gassian already in the first order in γ. The corresponding correction can be obtained from Eq. (16), taken only up to the first orded in nonlinearity:

$$\frac{\partial}{\partial\gamma}\left[\int \mathcal{D}x(\omega)\ p\left[y(\omega)|x(\omega)\right]\ p_x\left[x(\omega)\right]\right]\Bigg|_{\gamma=0} = 0 \quad (32)$$

Generally, such an integral would yield $p_x\left(x\right) = p_x^{(0)}\left(x\right)\left(1+\gamma p_x^{(0)}\left(x\right)\right)$, where $p_x^{(0)}$ is Gaussian, and $p_x^{(1)} \neq 0$. However, it is straightforward to show, that for the fiber optics channel described by Eq. (10), a Gaussian input distribution leads to non-Gaussian corrections in the output distribution starting only from the *second* order. Therefore, the requrement $p_y^{(1)} = 0$ is satisfied, when the input distribution is such that $p_x^{(1)} = 0$.

For the channel capacity, defined as the maximum value of the mutual information, we obtain:

$$C = W\log\left(1+\frac{P_0}{P_N}\right) - \Delta C_1 - \Delta C_2 + \mathcal{O}\left(\gamma^4\right) \quad (33)$$

The equation (33) yields the result for the fiber optics channel capacity in the second order in the nonlinearity γ. In the next section we will discuss the physical origins of the corrections ΔC_1 and ΔC_2.

VI. The Discussion and the Conclusions

In the spirit of the Shannon formula, the decrease of the capacity of a communication channel with a fixed bandwidth can be attributed to (i) the effective suppression of the signal power, and (ii) the enhancement of the noise. The corrections to the channel capacity, derived in the previous section, can be interpreted as resulting precisely from these two effects.

The four-wave scattering[2], induced by the fiber nonlinearity, inevitably leads to the processes, which generate photons with the frequencies outside the channel bandwidth. Such photons, are not recorded by the "receiver", and are lost for the purpose of the information transmission. This corresponds to an effective bandwidth power dissipation, and should therefore lead to a decrease of the channel capacity. Since for small nonlinearity this power loss $\Delta P \sim \gamma^2$, the dimension analysis implies

$$\Delta P = \frac{\gamma^2 P_0^2}{\alpha^2} F\left(\frac{\beta W^2}{\alpha}\right) \quad (34)$$

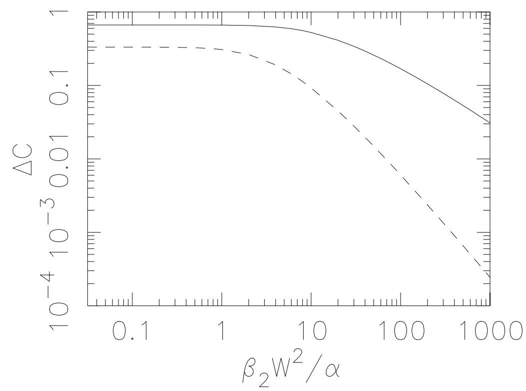

Fig. 2. The corrections to the channel capacity, ΔC_1 and ΔC_2, in units of $WN^2\gamma^2P_0^2\alpha^{-2}$, shown as functions of $|\beta_2|W^2/\alpha$ in the limit $\alpha d \gg 1$, $N \gg 1$. The correction ΔC_2 is represented by the solid line, while ΔC_1 corresponds to the dashed line. Note, that ΔC_1, which describes the effect of the power leakage from the bandwidth, is more strongly affected by the dispersion

Such scattering processes are suppressed, when the scattering leads to a subsantial change of the total momentum $\delta\kappa\left[\omega_1,\omega_2\to\omega_3,\omega\right]$, so that the corresponding scattering rate

$$S\left[\omega_1,\omega_2\to\omega_3,\omega\right] \sim \frac{\delta\left(\omega_1+\omega_2-\omega_3-\omega\right)}{1+\left(\delta\kappa/\kappa_0\right)^2} \quad (35)$$

In the spirit of the uncertainty realtion, $\kappa_0 \sim 1/L_{\rm eff}$, where $L_{\rm eff}$ corresponds to the length of the concentration of the power of the signal in the fiber. For a small absorption coefficient $\alpha \ll 1/d$ the distance $L_{\rm eff}$ is of the order of the fiber length d, while in the opposite limit $\alpha \gg 1$ the effective length $L_{\rm eff} \approx 1/\alpha$.

Using Eqn. (19), and the energy conservation $\omega_3 = \omega_1 + \omega_2 - \omega$, the momentum change $\delta\kappa\left[\omega_1,\omega_2\to\omega_3,\omega\right]$ can be expressed as

$$\delta\kappa = \beta_2\left(\omega-\omega_1\right)\left(\omega-\omega_1\right) \quad (36)$$

Substituting (36) into (35), for the channel capacity loss due to the bandwidth power "leakage", in the limit $P_0 \gg P_N$, and $\alpha d \gg 1$, we obtain

$$\begin{aligned}\Delta C_P &\sim W\frac{\Delta P}{P} \sim W\frac{\gamma^2P_0^2}{\alpha^2}\int_W d\omega_1\int_W d\omega_2\int_W d\omega_3 \\ &\times \int_{\omega\notin W} d\omega\ S\left[\omega_1,\omega_2\to\omega_3,\omega\right] \\ &= W\frac{\gamma^2P_0^2}{\alpha^2}\int_W d\omega_1\int_W d\omega_2\int_{\omega\notin W} d\omega \\ &\times \frac{1}{1+\left(\beta_2/\alpha\right)^2\left(\omega-\omega_1\right)^2\left(\omega-\omega_2\right)^2} \quad (37)\end{aligned}$$

which in the appropriate limit is consistent with ΔC_1.

In Fig. 2 we plot the dependence of ΔC_1 on the dimensionless parameter $\beta_2 W^2/\alpha$. Since momentum change $\delta\kappa$ is proportional to β_2, the increase of the dispersion leads to a strong suppression of the power leakage from the bandwidth window, and of the corresponding correction to the channel capacity.

In a communication system with many "fiber-amplifier" units, the fiber nonlinearity leads not only to the mixing of the signals at different frequencies, but also to the mixing of the signal with the noise. Qualitatively, this would correspond to an effective enhancement of the noise power in the system, and therefore to a loss of the channel capacity. This effect is not present, when the system has only one "fiber-amplifier" link, which explains the appearance of the $(N-1)$ factor in ΔC_2 and ΔC_3.

The effective noise enhancement is caused by the scattering processes, which involve a "signal photon" and a photon, produced due to spontaneous emission in one of the amplifiers. The total power of this extra noise can be expressed as

$$\begin{aligned} \frac{\Delta P_N}{P_0} &\sim \gamma^2 \frac{P_0 P_N}{\alpha^2} \int_W d\omega_1 \int_W d\omega_2 \int_W d\omega_3 \int_W d\omega \\ &\times S\left[\omega_1, \omega_2 \to \omega_3, \omega\right] \end{aligned} \tag{38}$$

The corresponding correction to the capacity

$$\begin{aligned} \Delta C_N &\sim W \frac{\Delta P_N}{P_N} \sim W \frac{\gamma^2 P_0^2}{\alpha^2} \int_W d\omega_1 \int_W d\omega_2 \int_W d\omega \\ &\times \frac{1}{1 + \left(\beta_2/\alpha\right)^2 \left(\omega - \omega_1\right)^2 \left(\omega - \omega_2\right)^2} \end{aligned} \tag{39}$$

where we assumed $\alpha d \gg 1$. In this limit (39) is up to a constant factor identical to ΔC_2.

The dependence of ΔC_2 on $\beta_2 W^2/\alpha$ is also shown in Fig. 2. Note, that ΔC_2 also decreases with the increase of the dispersion, but more slowly than ΔC_1. Since the scattering processes, which contribute to ΔC_1, need to "move" one of the frequencies out of the bandwidth window, they generally involve a substantial change of the total momentum, and are therefore more strongly affected by the dispersion.

The two physical effects, described above, determine the fundamental limit to the bit rate for a fiber optics communication system. As follows from our analysis (see Fig. 2), the relative contributions of ΔC_1 and ΔC_2, often referred to as the "four-wave mixing", can be suppressed by choosing a fiber with a large dispersion, or when using a larger bandwidth.

In our analysis, we treated the whole available bandwidth as a single channel. As a result, the cross-phase modulation[2], which severely limits the performance of advanced wavelength-division multiplexing systems[6] (WDM), does not affect the channel capacity. The reason for this seemingly contradictory behaviour, is that in a WDM system, the "receiver", tuned to a particular WDM channel, has no information on the signals at the other channels. Therefore, even in the absense of the "geniune" noise, the nonlinear interaction between different channels, leading to a change in the signal in any given channel, will be an effective noise source, thus limiting the communication rate. This limit however is not fundamental, and can be overcome by using the whole bandwidth all together.

In conclusion, we developed a perturbative method for the calculation of the channel capacity for fiber optics communication systems. We obtained analytical expressions for the corrections to the Shannon formula due to fiber nonlinearity. We have shown that, compared to the Shannon limit, the actual channel capacity is substantially suppressed by the photon scattering processes, caused by the fiber nonlinearity.

Appendix

I. Perturbative Solution of the Propagation Equation

In this Appendix we describe the perturbative solutuion of the nonlinear equation (10) with the boundary condition

$$A(0,t) = \int d\omega \; x(\omega) \exp(i\omega t) \tag{40}$$

We represent $A(z,t)$ as a power series

$$\begin{aligned} A(z,t) &= \int d\omega \exp\left(i\omega t - \left[\frac{\alpha}{2} + i\kappa_\omega\right] z\right) \sum_{n=0}^{\infty} \gamma^\ell \\ &\times \mathcal{F}_\ell(z,\omega) \end{aligned} \tag{41}$$

where κ_ω is defined is (19). Substituting (41) into Eqns. (10),(40), we obtain:
(i) for $\ell = 0$

$$\frac{\partial \mathcal{F}_0(z,\omega)}{\partial z} = 0 \tag{42}$$

$$\mathcal{F}_0(0,\omega) = x(\omega) \tag{43}$$

(ii) for $\ell \neq 0$

$$\begin{aligned} \frac{\partial \mathcal{F}_\ell(z,\omega)}{\partial z} &= i \sum_{\ell_1,\ell_2,\ell_3=1}^{\ell-1} \delta_{\ell-1,\ell_1+\ell_2+\ell_3} \int d\omega_1 \int d\omega_2 \\ &\times \mathcal{F}_{\ell_1}(z,\omega_1) \mathcal{F}_{\ell_2}(z,\omega_2) \mathcal{F}^*_{\ell_3}(z,\omega_1+\omega_1-\omega) \\ &\times \exp\left[i\left(\kappa_{\omega_1}+\kappa_{\omega_1}-\kappa_{\omega_1+\omega_2-\omega}-\kappa_\omega\right)\right] \end{aligned} \tag{44}$$

$$\mathcal{F}_\ell(0,\omega) = 0 \tag{45}$$

where δ in the Kronekker's delta-function.

For any ℓ, these equations reduce to *linear* first order differential equation, and can be solved straightforwardly. For example, the solutions for the first three terms in the anzats (41) are given by:

$$\mathcal{F}_0(z,\omega) = x(\omega) \tag{46}$$

$$\begin{aligned} \mathcal{F}_1(z,\omega) &= \int d\omega_1 \int d\omega_2 \; F^\omega_{\omega_1\omega_2} x(\omega_1) x(\omega_2) \\ &\times x^*(\omega_1+\omega_2-\omega) \end{aligned} \tag{47}$$

$$\begin{aligned} \mathcal{F}_2(z,\omega) &= \int d\omega_1 \int d\omega_2 \int d\omega_3 \int d\bar{\omega} \left[G^\omega_{\omega_1\omega_2\omega_3\bar{m}}\right. \\ &\times x^*(\omega_1) x^*(\omega_2) x(\omega_1+\omega_2-\bar{\omega}) x(\omega_3) \\ &\times x(\omega+\bar{\omega}-\omega_3) + H^\omega_{\omega_1\omega_2\omega_3\bar{\omega}} x(\omega_1) x(\omega_2) x(\omega_3) \\ &\times \left. x^*(\omega_1+\omega_2-\bar{\omega}) x^*(\bar{\omega}+\omega_3-\omega)\right] \end{aligned} \tag{48}$$

Here the functions F, G and H are given by

$$F^\omega_{\omega_1\omega_2} = i \frac{1 - \exp\left(-\alpha d - i\phi^\omega_{\omega_1\omega_2} d\right)}{\alpha + i\phi^\omega_{\omega_1\omega_2}}$$

$$
\begin{aligned}
G^{m}_{\omega_1\omega_2\omega_3\bar\omega} &= -\left(F^{\bar\omega}_{\omega_1\omega_2}\right)^* \left(F^{\omega_3}_{\omega\bar\omega}\right)^* \\
&\times \;\left(1-\exp\left(-\alpha d + i\phi^{\omega_3}_{\omega\bar\omega} d\right)\right) \\
&- \left(\frac{F^{\bar\omega}_{\omega_1\omega_2}}{\frac{1}{F^{\bar\omega}_{\omega_1\omega_2}}+\frac{1}{F^{\omega_3}_{\omega\bar\omega}}}\right)^* \\
&\times \;\left(1-\exp\left(-2\alpha d + \left(\phi^{\bar\omega}_{\omega_1\omega_2}+\phi^{\omega_3}_{\omega\bar\omega}\right) d\right)\right) \\
H^{\omega}_{\omega_1\omega_2\omega_3\bar\omega} &= 2\left(F^{\bar\omega}_{\omega_1\omega_2}\right)\left(F^{m}_{\bar\omega\omega_3}\right) \\
&\times \;\left(1-\exp\left(-\alpha d - i\phi^{\omega}_{\bar\omega\omega_3} d\right)\right) \\
&- 2\left(\frac{F^{\bar\omega}_{\omega_1\omega_2}}{\frac{1}{F^{\bar\omega}_{\omega_1\omega_2}}+\frac{1}{F^{\omega}_{\bar\omega\omega_3}}}\right)^* \\
&\times \;\left(1-\exp\left(-2\alpha d - i\left(\phi^{\bar\omega}_{\bar\omega_1\omega_2}+\phi^{\omega}_{\bar\omega\omega_3}\right) d\right)\right)
\end{aligned}
$$

where

$$\phi^{\omega}_{\omega_1\omega_2} = \left(\kappa_{\omega_1}+\kappa_{\omega_2}-\kappa_{\omega_1+\omega_2-\omega}-\kappa_{\omega}\right)$$

II. The Perturbative Calculation of the Channel Capacity

In this Appendix, we describe the perturbative calculation of the capacity of the simple nonlinear channel

$$y(\omega) = x(\omega)\exp\left(i\gamma\phi\left[x(\omega)\right]\right) + n(\omega) \tag{49}$$

Here $\phi[x]$ is an arbitrary real function, and the noise $n(\omega)$ is a Gaussian random variable:

$$p_n\left[n(\omega)\right] \sim \exp\left[-\frac{\left|n(\omega)\right|^2}{P_N}\right] \tag{50}$$

The fact, that the noise in this model is additive, implies that the conditional distribution $p(y|x)$ is fixed, and defined by the noise distribution p_n:

$$p(y \mid x) = p_n\left[y - x\exp\left(i\gamma\phi(x)\right)\right] \tag{51}$$

It is therefore straightforward to show, that the entropy $H[y|x]$ does not depend on γ:

$$
\begin{aligned}
H[y|x] &= -\int dy p_n\left(y - xe^{i\gamma\phi(x)}\right) \\
&\quad\times\log\left[p_n\left(y - xe^{i\gamma\phi(x)}\right)\right] \\
&= -\int dz p_n(z)\log\left[p_n(z)\right] \\
&= \log\left[2\pi e P_N\right]
\end{aligned} \tag{52}
$$

In order to calculate the entropy $H[y]$, we represent the output distribution as a power series in γ:

$$p_y(y) = p_y^0(y)\left[1+\sum_{n=1}^{\infty} p_y^{(n)}(y)\right] \tag{53}$$

where $p_y^0(y)$ is the “unperturbed”, Gaussian distribution

$$p_y^0(y) = \frac{1}{\pi\left(P_0+P_N\right)}\exp\left(-\frac{|y|^2}{P_0+P_N}\right) \tag{54}$$

corresponding to the linear channel $y = x + n$. Substituting (53) into the definition of the entropy H_y, Eq. (5), we obtain:

$$
\begin{aligned}
H_y &= \log\left[2\pi e\left(P_0+P_N\right)\right] - \frac{1}{2}\int p_y^{(1)}(y)^2 p_y^0(y) \\
&+ \frac{1}{P_0+P_N}\left[\int dy\,|y|^2 p_y(y)\right. \\
&- \left.\int dy\,|y|^2 p_y^0(y)\right]
\end{aligned} \tag{55}
$$

The second term in Eq. (55), $(1/2)\int p_y^{(1)}(y)^2 p_y^0(y)$, represents the difference of the output distribution from Gaussian, and corresponds to the contribution ΔH_y in Eq. (20). Note, that in the second order in nonlinearity the deviations of the output distribution from Gaussian lead to a *decrease* of capacity.

The third term, $(1/(P_0+P_N))\left[\int dy\,|y|^2 p_y(y) - \int dy\,|y|^2 p_y^0(y)\right]$, is proportional to the change of the output power, $\int dy\,|y|^2 p_y(y)$, due to nonlinearity, and corresponds to ΔC_1 in Eq. (20). Generally, the nonlinearity leads to energy exchange between different degrees of freedom in the channel (e.g. between different frequencies), and to the power leakage out of the bandwidth window. However, for the specific (nad non-generic) example, chosen in the present Appendix, this exchange is absent, since the output power

$$\left\langle |y|^2\right\rangle = \left\langle |x|^2\right\rangle + \left\langle |n|^2\right\rangle = P_0 + P_N \tag{56}$$

does not depend on the nonlinearity.

Substituting (56) in Eq. (55), and using Eq. (52), for the mutual information R we obtain:

$$R = \log\left[1+\frac{P_0}{P_N}\right] - \frac{1}{2}\int p_y^{(1)}(y)^2 p_y^0(y) \tag{57}$$

As immediately follows from Eq. (57), the channel capacity, equal to the maximum of the mutual information, is given by the Shannon formula (1), and is achieved when

$$p_y^{(1)}(y) = 0 \tag{58}$$

The next step is to calculate the *input* distribution

$$p_x(x) = p_x^0(x)\left[1+\sum_{n=1}^{\infty} p_x^{(n)}(x)\right] \tag{59}$$

corresponding to (58). The general relation between the input and the output distributions is defined by the conditional distribution $p(y|x)$:

$$p_y(y) = \int dx\; p(y|x)\;\; p_x(x) \tag{60}$$

and, considered as an equation for $p(x)$, is a Fredholm intergal equation of the first kind. Note however, that since Eq. (58) represents not the whole output distribution, but only it's first order term $p_y^{(1)}(y)$, we can expand Eq. (60)

and keep only the terms up to the first order in γ. We obtain:

$$\int dx \;\; p\left(y|x\right)|_{\gamma=0} \;\; p_x^0\left(x\right) \;\; p_x^{(1)}\left(x\right)$$
$$+ \;\; \int dx \;\; \frac{\partial}{\partial\gamma} p\left(y|x\right)\Big|_{\gamma=0} \;\; p_x^0\left(x\right) = 0, \qquad (61)$$

Substituting here the conditional distribution from Eq. (51), we obtain:

$$\int dx \; p_n\left(y-x\right) p_x^0\left(x\right) p_1\left(x\right) = \frac{i}{P_N}$$
$$\times \int dx \; p_n\left(y-x\right) p_x^0\left(x\right) \phi\left(x\right)\left(x^* y - y^* x\right), \qquad (62)$$

Using the identity

$$y p_n\left(y-x\right) = \left(x + P_N \frac{\partial}{\partial x^*}\right) p_n\left(y-x\right), \qquad (63)$$

and integrating by parts, we can represent the right hand side of (62) as follows:

$$i \int dx \; p_n\left(y-x\right) p_x^0\left(x\right) \frac{i}{P_N} \phi\left(x\right)\left(x^* y - y^* x\right) =$$
$$i \int dx \; p_n\left(y-x\right) p_x^0\left(x\right) \left(x^* \frac{\partial\phi\left(x\right)}{\partial x^*} - x \frac{\partial\phi\left(x\right)}{\partial x}\right) \qquad (64)$$

Therefore, as follows from Eqns. (62) and (64), the input distribution

$$p_x^{(1)}\left(x\right) = i\left(x^* \frac{\partial\phi\left(x\right)}{\partial x^*} - x \frac{\partial\phi\left(x\right)}{\partial x}\right) \qquad (65)$$

This procedure can be followed up for all orders in γ. By a direct calculation, it is straightforward to show, that the channel capacity is represented by the Shannon result (1), which is achieved when for any $n > 1$

$$\left\{ \begin{array}{l} p_x^{(n)}\left(x\right) = 0 \\ p_y^{(n)}\left(y\right) = 0 \end{array} \right. \qquad (66)$$

Subsitituting (66) and (65) into Eq. (59), for the input distibution we finally obtain:

$$p_x\left(x\right) = \frac{1}{\pi P_0}\left[1 + i\gamma\left(x^* \frac{\partial\phi\left(x\right)}{\partial x^*} - x \frac{\partial\phi\left(x\right)}{\partial x}\right)\right]$$
$$\times \;\; \exp\left[-\frac{\left|x\right|^2}{P_0}\right] \qquad (67)$$

with the corresponding channel capacity

$$C = W \log\left[1 + \frac{P_0}{P_N}\right] \qquad (68)$$

This result has a simple physical meaning. When the input distribution is organized in such a way, that the quantity $z = x \exp\left(i\gamma\phi\left(x\right)\right)$ has the Gaussian distribution, then, considering z as input, the communication channel becomes linear: $y = z + n$, and the channel capacity is therefore given by the Shannon formula (1),(68). The corresponding input distribution is then defined by the Jacobian of the transformation from $x \equiv x_R + i x_I$ to $z \equiv z_R + i z_I$, $\partial\left(z_R, z_I\right)/\partial\left(x_R, x_I\right)$ (note, that x_R, x_I, z_R, z_I are defined as *real* variables):

$$p_x\left(x\right) = \frac{1}{\pi P_0} \frac{\partial\left(z_R, z_I\right)}{\partial\left(x_R, x_I\right)} \exp\left[-\frac{\left|x_R^2 + x_I^2\right|^2}{P_0}\right] \qquad (69)$$

which reduces to the distribution (67), since

$$\frac{\partial\left(z_R, z_I\right)}{\partial\left(x_R, x_I\right)} = 1 + \gamma x_R \frac{\partial\phi\left(x_R, x_I\right)}{\partial x_I} - \gamma x_I \frac{\partial\phi\left(x_R, x_I\right)}{\partial x_R}$$
$$\equiv 1 + i\gamma x^* \frac{\partial\phi\left(x, x^*\right)}{\partial x^*} - i\gamma x \frac{\partial\phi\left(x, x^*\right)}{\partial x} \qquad (70)$$

This result should be contrasted to the so called "Gaussian estimate" of the channel capacity[7]. In the latter appoach, the information channel is described by the *joint Gaussian* distribution

$$\mathcal{P}\left(x_\omega, y_\omega\right) \sim \exp\left(-\left[x_\omega^* \;\; y_\omega^*\right] \mathcal{A} \left[\begin{array}{c} x_\omega \\ y_\omega \end{array}\right]\right) \qquad (71)$$

where

$$\mathcal{A} = \left[\begin{array}{cc} \langle x_\omega^* x_\omega \rangle & \langle x_\omega^* y_\omega \rangle \\ \langle y_\omega^* x_\omega \rangle & \langle y_\omega^* y_\omega \rangle \end{array}\right]^{-1} \qquad (72)$$

The channel capacity is then estimated as the mutual information, corresponding to the distribution (71):

$$C_G = \int d\omega \log\left[\frac{\langle x_\omega^* x_\omega \rangle \langle y_\omega^* y_\omega \rangle}{\langle x_\omega^* x_\omega \rangle \langle y_\omega^* y_\omega \rangle - \langle x_\omega^* y_\omega \rangle \langle y_\omega^* x_\omega \rangle}\right] \qquad (73)$$

Under the constraint of the fixed input power $\int d\omega \langle |x_\omega|^2 \rangle$, the estimate (73) was shown[7] to give the low bound to the channel capacity.

For the model channel considered in the present Appendix, the "Gaussian estimate" yields an expression, *different* from the Shannon result. For example, when $\phi\left(x\right) = |x|^2$, we obtain

$$C_G = -W \log\left[1 - \frac{P_0}{\left(P_0 + P_N\right)\left(1 + \gamma^2 P_0^2\right)^2}\right]$$
$$= W \log\left[1 + \frac{P_0}{P_N}\right] - 2W\gamma^2 P_0^2 \frac{P_0}{P_N} + \mathcal{O}\left(\gamma^4\right) \qquad (74)$$

which, as expected, is *smaller* that the actual channel capacity (68). Note, that the difference between the exact channel capacity and the Gaussian estimate

$$\delta C \equiv C - C_G = W \log\left[1 + \frac{P_0}{P_N}\left(1 - \frac{1}{\left(1 + \gamma^2 P_0^2\right)^2}\right)\right]$$
$$= 2W\gamma^2 P_0^2 \frac{P_0}{P_N} + \mathcal{O}\left(\gamma^4\right) \qquad (75)$$

is not merely a constant scale factor, but a nontrivial function of the signal to noise ratio, and the nonlinerity.

Even when the input distribution *is Gaussian*, like e.g. when the phase ϕ depends on x via the "power" $|x|^2$, the Gaussian Estimate does not yield the exact result. The reason for this behaviour is that the joint Gaussian distribution does not correctly reproduce the *conditional* distribution $p(y \mid x)$.

For an essentially noinlinear system (e.g. a fiber optics communication channel), there is generally very little *apriori* knowledge about the parametric dependence of the Channel Capacity on the signal to noise ratio and other system parameters. In this case, the Gaussian Estimate for the channel capacity can be (should be?) viewed as a very unreliable method, as there is no way to separate it's artefacts from the actual behaviour of the channel capacity.

References

[1] C. E. Shannon, *A Mathematical Theory of Communication.* The Bell System Technical Journal vol 27, pp. 379 - 423, 623-656. 1948.

[2] G.P.Agrawal, *Nonlinear Fiber Optics*, Academic Press, San Diego, 1995.

[3] R.K.Dodd, J.C.Eilbeck, J.D.Gibbon, and H.C.Morris, *Solitons and Nonlinear Wave Equations*, Academic Press, New York, 1984.

[4] G. L. Lamb, Jr., *Elements of Soliton Theory*, Wiley, New York, 1980.

[5] J. P. Gordon and L. F. Mollenauer, *Phase noise in photonic communication systems using linear amplifiers.* Optics Letters, vol. 23, pp. 1351-1353. Dec. 1990.

[6] G. P. Agrawal, *Fiber-Optic Communication Systems*, John Wiley and Sons, New York, 1997.

[7] I. Teletar, unpublished.